Forecast Standard Errors

- Wooldridge, Chapter 6.4
- Multiple Regression

$$y_{t+h} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + e_t$$

- Includes intercept, trend, and autoregressive models (x can be lagged y)
- OLS estimate

$$y_{t+h} = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \cdots + \hat{\beta}_k x_{kt} + \hat{e}_t$$

- Point forecast

$$\hat{y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 x_{1T} + \hat{\beta}_2 x_{2T} + \cdots + \hat{\beta}_k x_{kT}$$

Forecast Error

- Forecast error

$$\hat{e}_{T+h} = y_{T+h} - \hat{y}_{T+h}$$

- Variance of forecast error

$$\text{var}(\hat{e}_{T+h}) = \text{var}(y_{T+h}) + \text{var}(\hat{y}_{T+h})$$

$$= \sigma^2 + \text{var}(\hat{y}_{T+h})$$

- Two components:
  - Equation (model) variance σ^2
  - Estimation variance $\text{var}(\hat{y}_{T+h})$

Equation Variance

- Variance $\sigma^2 = \text{var}(y_{t+h}) = \text{var}(e_t)$
- Residuals $\hat{e}_t = y_{t+h} - \hat{y}_{t+h}$
- Residual variance $$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_t^2$$
- Residual standard deviation $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$
- In Stata, displayed as “Root MSE”, or
  - `gen s = e(rmse)`

Estimation Variance

- Point prediction

$$\hat{y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 x_{1T} + \hat{\beta}_2 x_{2T} + \cdots + \hat{\beta}_k x_{kT}$$

- This is an estimate of the regression function at these values of the x's

$$\text{var}\left(\hat{y}_{T+h}\right) = \text{var}\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1T} + \hat{\beta}_2 x_{2T} + \cdots + \hat{\beta}_k x_{kT}\right)$$

- This is a function of the variances of the OLS estimates, weighted by the x's

Estimation Standard Errors

- Estimation standard error

$$se(\hat{y}_{T+h}) = \sqrt{\operatorname{var}(\hat{y}_{T+h})}$$

- These are standard errors for the conditional mean, not for forecasts
- Computed in Stata using *stdp* option for *predict* command
  – `predict s, stdp`
- Important: This is different than *stdf*

Forecast standard error

$$\operatorname{var}\left(\hat{e}_{T+h}\right) = \sigma^2 + \operatorname{var}\left(\hat{y}_{T+h}\right)$$

$$se\left(\hat{e}_{T+h}\right) = \sqrt{\hat{\sigma}^2 + \operatorname{var}\left(\hat{y}_{T+h}\right)}$$

$$= \sqrt{\hat{\sigma}^2 + se\left(\hat{y}_{T+h}\right)^2}$$

- Computed in Stata using stdf option
  - `predict s, stdf`
- Typically will be close to (just a little larger than) $\hat{\sigma}$

GDP Example

```
. regress gdp L.gdp

      Source |       SS           df       MS      Number of obs   =       278
-------------+----------------------------------   F(1, 276)       =     44.32
       Model |  585.172068         1  585.172068   Prob > F        =    0.0000
    Residual |  3643.98246       276   13.202835   R-squared       =    0.1384
-------------+----------------------------------   Adj R-squared   =    0.1352
       Total |  4229.15453       277  15.2677059   Root MSE        =    3.6336

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |
         L1. |   .3714752   .0557984     6.66   0.000     .2616307    .4813197
             |
       _cons |   2.036509   .2826904     7.20   0.000     1.480006    2.593013
------------------------------------------------------------------------------
```

- The equation standard error is “Root MSE”
  - 3.63

GDP Example

	time	s	sp	sf
280.	2017q1	3.633571	.2301621	3.640853

- Notice
  - s equals "Root MSE" from regression output
  - The estimates satisfy the relationship

$$sf^2 = \sqrt{sp^2 + s^2}$$

  - sf and s are very close
  - sf (standard error of forecast) is better
    - But s (error standard deviation) is often sufficient

Two-Step-Ahead Point Forecasting

- Three methods
  - Plug-in
    - Calculates optimal forecast as function of AR model
    - Replaces unknowns with estimates
  - Iterated
    - Calculates one-step forecast, and then iterates to get second-step forecast
  - Direct
    - Estimates 2-step regression function, and uses this for forecast
- We start with point forecasts, and then discuss interval forecasts

Plug-In Method

- By back-substitution

$$
\begin{aligned}
y_t &= \alpha + \beta y_{t-1} + e_t \\
&= \alpha + \beta\left(\alpha + \beta y_{t-2} + e_{t-1}\right) + e_t \\
&= \left(1 + \beta\right)\alpha + \beta^2 y_{t-2} + e_t + \beta e_{t-1}
\end{aligned}
$$

- Thus

$$
y_{T+2} = \left(1 + \beta\right)\alpha + \beta^2 y_T + e_{T+2} + \beta e_{T+1}
$$

$$
E\left(y_{T+2} \mid \Omega_T\right) = \left(1 + \beta\right)\alpha + \beta^2 y_T
$$

Point Forecast

- The optimal forecast is

$$\hat{y}_{T+2|T} = (1+\beta)\alpha + \beta^2 y_T$$

- This is a function of the AR(1) parameters
- Plug-in (replace unknowns with estimates) to obtain a feasible forecast

$$\hat{y}_{T+2|T} = (1+\hat{\beta})\hat{\alpha} + \hat{\beta}^2 y_T$$

- This method is feasible but cumbersome for multi-step forecasts and complicated models

Iterated Method

- Take conditional expectations at time T

$$y_{T+2} = \alpha + \beta y_{T+1} + e_{T+2}$$

$$E(y_{T+2} \mid \Omega_T) = \alpha + \beta E(y_{T+1} \mid \Omega_T) + E(e_{T+2} \mid \Omega_T)$$

$$= \alpha + \beta E(y_{T+1} \mid \Omega_T)$$

- The left-side is the 2-step forecast, the right-side is linear in the 1-step forecast. Thus:

$$\hat{y}_{T+2|T} = \alpha + \beta \hat{y}_{T+1|T}$$

Iteration

- We already know how to compute the one-step point forecast

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\beta} y_T$$

- The second step iterates on the one-step

$$\hat{y}_{T+2|T} = \hat{\alpha} + \hat{\beta} \hat{y}_{T+1|T}$$

- This method is convenient in linear models (our main focus)
- It does not work in nonlinear models
- It is less useful in regression contexts (later sections)

Direct Method

- We showed that

$$y_t = (1+\beta)\alpha + \beta^2 y_{t-2} + e_t + \beta e_{t-1}$$
$$= \alpha^* + \beta^* y_{t-2} + u_t$$

where

$$\alpha^* = (1+\beta)\alpha$$
$$\beta^* = \beta^2$$
$$u_t = e_t + \beta e_{t-1}$$

Estimation of Direct Method

- This is a regression

$$y_t = \alpha^* + \beta^* y_{t-2} + u_t$$

- The error is the two-step forecast error
- It can be estimated **directly** by least-squares
- This is actually different than the iterated estimator.
- The error *u* is not white noise, but is uncorrelated with the regressor

Example – GDP Growth

- Calculate 2-step forecast using plug-in method

$$\hat{y}_{T+2|T} = \left(1+\hat{\beta}\right)\hat{\alpha} + \hat{\beta}^2 y_T$$

- α=2.04, β=0.37, y_T=1.9

$$\begin{aligned}\hat{y}_{T+2|T} &= \left(1+\hat{\beta}\right)\times\hat{\alpha} + \hat{\beta}^2 y_T \\ &= \left(1+.37\right)\times 2.04 + .37^2 \times 1.9 \\ &= 3.1\%\end{aligned}$$

Example – GDP Growth

- Calculate 2-step forecast by iterated method

$$\hat{y}_{T+1|T} = \alpha + \beta\hat{y}_{|T}$$
$$\hat{y}_{T+2|T} = \alpha + \beta\hat{y}_{T+1|T}$$

- α=2.04, β=0.37, y_T=1.9
- One-step

$$\begin{aligned}\hat{y}_{T+1|T} &= \hat{\alpha} + \hat{\beta}\hat{y}_T \\ &= 2.04 + .37 \times 1.9 \\ &= 2.7\%\end{aligned}$$

- $y_{T+1|T}$=2.7
  - (also calculated in last class)

Example – GDP Growth

- Calculate 2-step forecast by iterated method

$$\hat{y}_{T+2|T} = \alpha + \beta \hat{y}_{T+1|T}$$

- α=2.04, β=0.37, $y_{T+1|T}$=2.7
- 2-step:

$$\begin{aligned} \hat{y}_{T+2|T} &= \hat{\alpha} + \hat{\beta} \hat{y}_{T+1|T} \\ &= 2.04 + .37 \times 2.7 \\ &= 3.1\% \end{aligned}$$

- Identical to plug-in forecast
  - They should be identical, up to rounding errors

Stata Forecast Command

```
. tsappend, add(2)

. forecast create ar1
  Forecast model ar1 started.

. estimates store model1

. forecast estimates model1
  Added estimation results from regress.
  Forecast model ar1 now contains 1 endogenous variable.

. forecast solve
```

- “forecast create [name1]”
- “estimates store [name2]” (after a regression)
- “forecast estimates [name2]” tells STATA to forecast using the estimates from *name2*
- “forecast solve” creates the forecasts, and stores them in the dataset with the name *f_[name]* where *name* is the variable name, e.g. *f_gdp*

STATA Forecast output

```
. list time f_gdp if time>tq(2016q4)
```

	time	f_gdp
280.	2017q1	2.7423122
281.	2017q2	3.0552104

- These are the one-step and two-step iterated point forecasts from the AR(1) model
- They are identical to what we calculated by hand

GDP Growth, Direct 2-step

```
. regress gdp L2.gdp

      Source |       SS           df       MS      Number of obs   =       277
-------------+----------------------------------   F(1, 275)       =     14.07
       Model |  205.170898         1  205.170898   Prob > F        =    0.0002
    Residual |  4010.72087       275  14.5844395   R-squared       =    0.0487
-------------+----------------------------------   Adj R-squared   =    0.0452
       Total |  4215.89177       276  15.2749702   Root MSE        =     3.819

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |
         L2. |    .219963   .0586458     3.75   0.000     .1045113    .3354147
             |
       _cons |   2.538776   .2973963     8.54   0.000     1.953314    3.124239
------------------------------------------------------------------------------
```

- Estimate

$$y_t = 2.54 + 0.22 y_{t-2} + \hat{u}_t$$

- Notice .22>.14=$.37^2$ from iterated

Direct 2-step-ahead Forecast

- 2-step forecast

$$\begin{aligned}\hat{y}_{T+2|T} &= \hat{\alpha}^{*} + \hat{\beta}^{*} y_T \\ &= 2.54 + 0.22 \times 1.9 \\ &= 3.0\%\end{aligned}$$

- It is very close to the iterated forecast, but this does not always happen.

2-Step Forecast Error

- Recall $y_t = \alpha^* + \beta^* y_{t-2} + u_t$

  where $u_t = e_t + \beta e_{t-1}$
- The equation error is *u*, not *e*
- It has variance

$$\begin{aligned} \mathrm{var}(u_t) &= \sigma_u^2 \\ &= \mathrm{var}(e_t + \beta e_{t-1}) \\ &= \left(1 + \beta^2\right)\sigma^2 \end{aligned}$$

- This is different than the one-step variance

Forecast variance estimation

- For forecast intervals, we need an estimate of

$$\mathrm{var}(u_t) = \sigma_u^2$$

- Not

$$\mathrm{var}(e_t) = \sigma^2$$

Plug-in Forecast variance estimation

- Use formula, and replace by estimates

$$\hat{\sigma}_u^2 = \left(1 + \hat{\beta}^2\right)\hat{\sigma}^2$$

$$\hat{\sigma}_u = \sqrt{\hat{\sigma}_u^2}$$

- This formula is hard to generalize beyond AR(1)

Example: GDP Growth Plug-in Estimate

- β=.37, σ=3.63

$$
\begin{aligned}
\hat{\sigma}_u &= \sqrt{\left(1+\hat{\beta}^2\right)\hat{\sigma}^2} \\
&= \sqrt{\left(1+.37^2\right)3.63^2} \\
&= 3.9
\end{aligned}
$$

Direct Forecast variance estimation

$$\hat{u}_t = y_t - \hat{\alpha}^* - \hat{\beta}^* y_{t-2}$$

$$\hat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2$$

Direct Estimate

```
. reg gdp L2.gdp

      Source |       SS           df       MS      Number of obs   =       277
-------------+----------------------------------   F(1, 275)       =     14.07
       Model |  205.170898         1  205.170898   Prob > F        =    0.0002
    Residual |  4010.72087       275  14.5844395   R-squared       =    0.0487
-------------+----------------------------------   Adj R-squared   =    0.0452
       Total |  4215.89177       276  15.2749702   Root MSE        =     3.819

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |
         L2. |    .219963   .0586458     3.75   0.000     .1045113    .3354147
             |
       _cons |   2.538776   .2973963     8.54   0.000     1.953314    3.124239
------------------------------------------------------------------------------
```

- Estimate $\hat{\sigma} = 3.82$
- Stdf $se(\hat{e}) = 3.82$

Iterated Forecast Variance Estimation

- Not easy to calculate directly
- The forecast errors *u* not a direct output
- Instead, it is typical to use *simulation* to calculate forecast variance
- This can be more flexible than the formulae
- Can be done in STATA using*forecast* command

```
. forecast solve, simulate(errors,statistic(stddev,prefix(sd_)) reps(1000))
```

Iterated Forecast Variance Estimation

```
. forecast solve, simulate(errors,statistic(stddev,prefix(sd_)) reps(1000))
```

- The *simulate* option creates simulated out-of-sample series from the model
- The *statistic* option tells STATA what to save (standard deviations)
- The *prefix* option tells STATA to save the standard deviations in the format *sd_name*, where "name" was the variable you are forecasting.
- The *reps* option tells STATA to use 1000 simulations (otherwise 50 is the default)
- This command creates the point forecasts *f_gdp* and standard derivations *sd_gdp*

GDP Example

- This shows the 1-step and 2-step point forecasts (2.7 and 3.1), and the 1-step and 2-step forecast standard errors (3.6 and 3.8)
- This is slightly smaller than the plug-in estimate 3.9

```
. list time f_gdp sd_gdp if time>tq(2016q4)
```

	time	f_gdp	sd_gdp
280.	2017q1	2.7423122	3.6455067
281.	2017q2	3.0552104	3.7546381

Two-Step-Ahead Intervals

- Normal Method
  - Forecast interval is point estimate, plus and minus the estimated standard deviation multiplied by a normal quantile
  - For a 95% interval:

$$\hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot z_{.025} = \hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot 1.96$$

  - For a 90% interval

$$\hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot z_{.05} = \hat{y}_{T+2|T} \pm \hat{\sigma}_u \cdot 1.645$$

GDP Growth 2-step intervals

- Plug-in
  - $y_{T+2|T}$=3.1%, σ_u=3.9
  - 3.1% ± 1.645*3.9 = [-3.3%, 9.5%]
- Iterated
  - $y_{T+2|T}$=3.1%, σ_u=3.8
  - 3.1% ± 1.645*3.8 = [-3.2%, 9.4%]
- Direct
  - $y_{T+2|T}$=3.0%, σ_u=3.8
  - 3.0% ± 1.645*3.8 = [-3.3%, 9.3%]

h-Step-Ahead Forecasting

$$\hat{y}_{T+h|T}$$

h-Step-Ahead back substitution

$$
\begin{aligned}
y_t &= \alpha + \beta y_{t-1} + e_t \\
&= \alpha + \beta\left(\alpha + \beta y_{t-2} + e_{t-1}\right) + e_t \\
&= \left(1 + \beta\right)\alpha + \beta^2\left(\alpha + \beta y_{t-3} + e_{t-2}\right) + e_t + \beta e_{t-1} \\
&= \left(1 + \beta + \beta^2\right)\alpha + \beta^3 y_{t-3} + e_t + \beta e_{t-1} + \beta^2 e_{t-2} \\
&= \left(1 + \beta + \beta^2 + \cdots + \beta^h\right)\alpha + \beta^h y_{t-h} + u_t \\
u_t &= e_t + \beta e_{t-1} + \beta^2 e_{t-2} + \cdots + \beta^{h-1} e_{t-h+1}
\end{aligned}
$$

h-Step-Ahead Point Forecast

- Optimal

$$E\left(y_{T+h} \mid \Omega_T\right) = \left(1 + \beta + \beta^2 + \cdots + \beta^h\right)\alpha + \beta^h y_T$$

- Plug-In

$$\hat{y}_{T+h|T} = \left(1 + \hat{\beta} + \hat{\beta}^2 + \cdots + \hat{\beta}^h\right)\hat{\alpha} + \hat{\beta}^h y_T$$

- Iterated

$$y_{T+h} = \alpha + \beta y_{T+h-1} + e_{T+h}$$

$$E(y_{T+h} \mid \Omega_T) = \alpha + \beta E(y_{T+h-1} \mid \Omega_T)$$

$$\hat{y}_{T+h|T} = \hat{\alpha} + \hat{\beta}\hat{y}_{T+h-1|T}$$

Direct Method

- Best Linear predictor

$$y_t = \alpha^* + \beta^* y_{t-h} + u_t$$

- Least-Squares estimator

$$y_t = \hat{\alpha}^* + \hat{\beta}^* y_{t-h} + \hat{u}_t$$

- h-step forecast

$$\hat{y}_{T+h|T} = \hat{\alpha}^* + \hat{\beta}^* y_T$$

4-Step Direct Estimation

```
use realgdpgrowth.dta
tsappend, add(4)
reg gdp L.gdp
reg gdp L2.gdp
reg gdp L3.gdp
reg gdp L4.gdp
```

- Four regressions, one for each forecast horizon

Direct Estimates

- Least Squares

$$y_t = 2.04 + 0.37 y_{t-1} + \hat{e}_t$$

$$y_t = 2.54 + 0.22 y_{t-2} + \hat{u}_t$$

$$y_t = 3.14 + 0.03 y_{t-3} + \hat{u}_t$$

$$y_t = 3.41 - 0.06 y_{t-4} + \hat{u}_t$$

4-Step Direct Point Forecast

```
reg gdp L.gdp
predict y1
reg gdp L2.gdp
predict y2
reg gdp L3.gdp
predict y3
reg gdp L4.gdp
predict y4
```

- There are 4 periods out-of-sample
- The **predict** command computes fitted values for observations which have the needed variables.
- For the regression on the first lag (L.gdp), this works only for the first out-of-sample observation, the remainder are coded as missing.
- For the regression on the second lag (L2.gdp), this works for the fist two out-of-sample observations

Forecasts

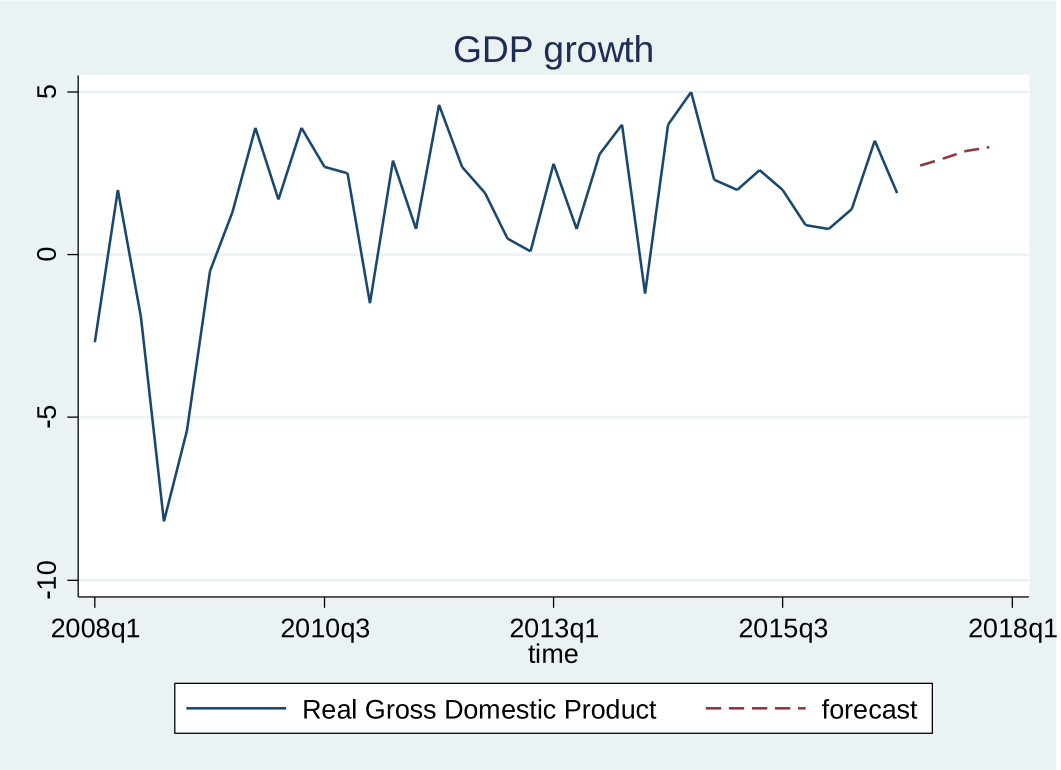
	time	y1	y2	y3	y4	p
279.	2016q4	3.336673	2.846725	3.164464	3.360592	.
280.	2017q1	2.742312	3.308647	3.182345	3.366302	2.742312
281.	2017q2	.	2.956706	3.244925	3.332045	2.956706
282.	2017q3	.	.	3.197245	3.212148	3.197245
283.	2017q4	.	.	.	3.303498	3.303498

- The direct forecast for each quarter is the final entry in each column, collected in "p"

4-Step Direct Point Forecast

```
egen p=rowfirst(y1 y2 y3 y4) if t>=tq(2017q1)
label variable p "forecast"
tsline gdp p if t>=tq(2008q1), title(GDP growth) lpattern (solid dash)
```

- The egen command is used in Stata for more complicated versions of "generate"
- `egen p=rowfirst(y1 y2 y3 y4)` takes the first variable in the list which is not missing



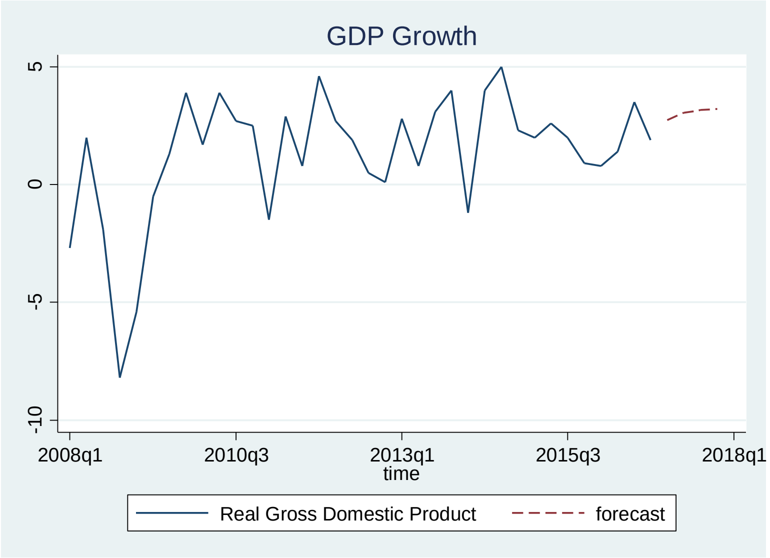
4-Step Iterated Point Forecast

```
use realgdpgrowth.dta
tsappend, add(4)
reg gdp L.gdp
forecast create ar1
estimate store model1
forecast estimates model1
forecast solve
gen pf = f_gdp if t>=tq(2017q1)
label variable pf “forecast”
tsline gdp pf if t>=tq(2008q1), title(GDP growth)
  lpattern (solid dash)
list time p pf if time>=tq(2017q1)
```

Direct and Iterated Point Estimates

	time	p	pf
280.	2017q1	2.742312	2.742312
281.	2017q2	2.956706	3.05521
282.	2017q3	3.197245	3.171444
283.	2017q4	3.303498	3.214622

- *p* is direct forecast
- *pf* is iterated forecast
- Very similar



Assignments

- Read Diebold through Chapter 7
- Problem Set # 6
  - Due Tuesday (2/28)
- Read Chapter 5 from *The Signal and the Noise*
  - Reading Reflection
  - Due Thursday (2/23)